Interviewee: CRU_04

Name: Phil Jones

Role / Title: Professor, Director CRU

Organisation: Climatic Research Unit, UEA

Date: 25th April 2014

Interviewer: Paula Goodale

==========

Q: Okay.

A: So the Met Office will probably try and get you to go to certain people. You've heard about Philip Brohan

Q: Yes, and we've had a provisional okay from Philip, although I believe he's abroad at the moment.

A: Yeah, I think he's in Mozambique at the moment. The other person on that citizen science / data rescue angle is someone called Rob Allan.

Q: Yes, we've also provisionally been in contact with Rob.

A: And he's at the same meeting as well I think.

Q: Do you know when they're back? Is that a longish--,

A: Well, I think it's only a week.

Q: All right, okay, so by the time we're back--,

A: They're just going for a sort of three or four day meeting and they'll probably have another--, maybe over the weekend, but they'll be back next week I think. Then the person we deal with mostly down there and who produces the various datasets that we'll talk about, is Colin Morice, M-o-r-i-c-e.

Q: That's an unusual spelling. So he's kind of responsible for the Hadley side of the datasets.

A: Hadley side of things, yes, and he's relatively young, so he's only been doing it a few years. If you look at the scientific papers, two papers came out in 2012 and they're in the *Journal of Geophysical Research*. So on those issues they tend to publish in American journals, to get them more widely read, rather than Royal Met Soc journals. That's partly because those journals have a higher impact factor and that sort of issue. The other guy, the fourth one, is John Kennedy, and he's the one who does all the marine stuff. Although Philip and Rob are also on the digitising of data rescue, there's a lot of marine stuff as well as land stuff.

Q: Yeah, I had a very illuminating conversation with Clive yesterday on that side of things, so it was useful to get that kind of starting point for that area of work. Okay, well that's really useful, thank you.

A: Yeah. If you just mention one other thing when you come to questions, it might be worth near the end if there's a bit of time, to mention something called reanalysis.

Q: Yes, and again that's something I did want to ask you about.

A: Okay, fine, we'll come onto that.

Q: And some of the other guys have mentioned as well, yes.

A: We'll come to that, because that's what a lot of the digitising and rescue is leading towards.

Q: Yeah. In fact, let me make sure I've got the note there. [Pause]

Q: Okay. No, that's great. So should we start with your own work at CRU and, you know, how that's developed and changed over time, a little bit of a potted history.

A: Okay. I've been here since November 1976. I came on a three-year research contract, which was not too dissimilar from some of the things I do now. So at the time, I had finished a PhD on a sort of related topic, in fact it was on a catchment model of the River Tyne. My aim at the end of that was to get a job in the water industry, one of the water companies or the Institute of Hydrology as it was then.

Q: So were you a geographer by trade?

A: No. I did an Environmental Science, first degree, at Lancaster because they accepted me, and this place didn't. After that, I got more interested in the water side of the environment and went on to do a PhD, well a Master's first in hydrology and then a PhD on this catchment model of the River Tyne. The aim was then to get into the water industry, as I said. But at that time I needed job at the end of the PhD and here was a job which I had managed to get, in '76. It was partly because the water industry was going through the drought, the '75/'76 drought, and they weren't doing much hiring at the time, there weren't many jobs, and this was a job, and in retrospect, 37 or 38 years later, it was quite a good job to have got. But, it involved a number of short-term research projects, a bit like what David and Harry are on now in terms of they're not funded by the university, I don't know whether you got into that.

Q: A little bit, yeah.

A: Funded by research, soft money research basically, a bit like you are, maybe.

Q: Yes, exactly.

A: So I was on that soft money research projects, mainly. There was one NERC project, which was back doing a bit more like my PhD for a while. But after about 1984, I got onto a source of money funded by the US Department of Energy. And we've had 10—we've had 33 years of three-year contracts, with small hiatuses every now and then of a few months between one contract and the other, so we've had pretty continuous funding.

Q: From the US?

A: From the US. So that's what's funding David at the moment, and it has funded others in the unit over the years.

Q: And that's largely government money, is it?

A: It's largely US money, yeah. I don't think we would have got that from UK-based sources, unless you went to work at the Hadley Centre where they have a big block grant of millions.

Q: But that came later, didn't it?

A: That sort of came later, when Margaret Thatcher set up the Hadley Centre in the late '80s, early '90s, I think it was about '88/'89. We kept on working independently of the Met Office, and then working more and more with them over the years. So, in terms of the datasets that you've come down for, the first paper was one in '82. Producing the global temperature record was we just thought at the time an interesting thing to do. We were used to using datasets, mainly pressure data, to look at variability in climate over the years and we thought it would be good get a temperature one as well as pressure. So the first paper was in '82, and we've revisited the whole thing several times, in '86, again in about '94, and then another one in 2003, 2006 with Philip Brohan, and 2012 with Colin Morice. So there've been various versions.

Q: Are they updates of the same?

A: They're updates, but also extensions and improvements in terms of getting more station data, and using improved ways of analysing the data as time progressed.

Q: Yeah, and that's knowledge that you've kind of developed and applied.

A: Yeah, and a lot of it involved better data and assessing more data. So when we first did it in '82, we managed to get access to a magnetic tape, data that was digitised in the US from something called World Weather Records, and I'll show you some of those in the library.

Q: Yeah, that'll be interesting.

A: We weren't able to look at the quality of the data then, we were just able to use it, so we had to screen it for bad values that been digitised incorrectly or just printing mistakes in the publications.

Q: That was done by eye, was it?

A: We check for outliers, I think David mentioned this. You may be looking at data by sight initially, and you've got 12 months' values for each of the strings of years, so basically you were just looking at a period. Our base periods have changed over the years. We started off with '51 to '70, '51 to '80, and now we're on '61 to '90, and--,

Q: Yeah, and you're comparing with those?

A: Comparing with those. If you just take the average value for January for '61 to '90 and you work out the standard deviation for those 30 values, then you can flag numbers that are more than a number of standard deviations from the mean. That was a way eliminating bad values

which were four or five standard deviations away from the mean, so that works generally quite well.

Q: And you just discount those or would you try and--,

A: We just remove those, except where there were one or two issues where you could really figure out what to do with them. In the original piece of work, I just remembered one thing, for some reason all the Greenland values, years ago for November in this original dataset, were all instead of being negative, they were all positive.

Q: Right. Oh, yeah, David told me about this, yeah.

A: So we just switched the sign on them and they were all fine then. So in '86 we then moved to assessing the quality of the data a bit more, computing had improved by then, even from '82. Coming from Newcastle, moving here in '76, was a backwards move for my computing, because in my PhD at Newcastle they'd had a massive IBM machine, one of the best ones at the time, so you could do lots of things. When I came down here there was this ICL system called George, which was a backward step. And so a lot of the initial work in the '80s was done at Cambridge, we used to submit cards and they went to run there because the computer here wasn't good enough to do the work. But by the mid-'80s we had a better computer here, we had a DEC system using VAX.

Q: Yes, I remember them.

A: And so we were able to then assess the quality of the data, and in that, we produced several books. [Sound of papers being moved about]

Q: It's a bit like a telephone directory.

A: Yeah, but this is how the US Dept of Energy produced their reports. There was probably on the order of about 100 of these, not all ours. One of them was just looking at data sources around the world. This was done with a couple of colleagues in America. Some of the money from America went through these other institutions there, but we got our own grants for the last 30 years. These were just lists of stations, and then lists of sources of historical data which we trawled from the Met Office library.

Q: So it is kind of like a directory.

A: Yeah. Then we're just listing stations. Here's one, Montrose in Scotland, and the years of record and the sources and the locations. The interesting thing about the locations now is that, in the past we didn't really worry about getting the exact location, so things were given to a tenth of a degree of latitude and longitude. Now people want to try and see these sites with Google Earth and so if you just put these numbers in you'd probably find this station's in the sea, for that the location.

Q: [Laughs] So, it's because it's not precise enough, no.

A: It's not precise enough. A lot of met services still haven't got better accuracy, because you need about four decimal places in this system to be able to find it.

Q: I've seen on the Met Office, they've got a website now, you can look at individual stations.

A: Yeah. They've done all the ones in the UK, and a few other countries have and sent more details, but they're the countries you expect to do it and a lot of other countries haven't. So that was one report. Then another one was then looking at the various stations in the northern hemisphere, there's also a smaller group of them in the southern hemisphere as well, and what we were doing was assessing the quality of the records.

Q: Okay. So there you've got the--,

A: We've got the temperatures here, so these are the temperatures at Aberdeen and another place in Scotland. So what you do-- this is all done more automated now and there are lots of different ways of doing this now, but then we were just doing this all by just plotting diagrams like this. It wasn't until they got a better printer here and one of those old sort of--,

Q: Yeah, the dot matrix?

A: No, not the dot matrix ones, the sort of graphics plots that runs on a drum that go round,

Q: Yes, yes, not seen one of those for a long time.

A: No, I don't think they've probably got them now. But if you plot the annual values at Aberdeen as a difference from those at Gordon Castle, or vice versa, then what you should get is that the temperatures should vary about a straight horizontal line, if they've both measuring consistently. Now if one of them moves, then you might get a jump to a new level, so with two sites, you don't know whether it's Aberdeen or Gordon Castle. But if you look at other sites near Aberdeen. The initial Aberdeen record when it began in the 19th century was in the town itself and then in the late '40s it moved to the airport at Dyce.

Q: Right, and that changed the temperature.

A: That's what caused it, yeah, it got cooler. So because you're plotting Gordon Castle minus Aberdeen, that actually implies, it looks as though Gordon Castle has got warmer, but in fact Gordon Castle's the same the whole time, but Aberdeen's got cooler 'cause it moved out of the city to a cooler area, inland. The city site at Aberdeen is warmer because it's on the coast basically, and also Dyce is probably a little bit higher as well. So there's another example there, and then there's an example here for two places in India, the same type of thing. [Sound of pages being turned] There's two in Japan, and really, the jump is just depending which way round you difference them.

Q: So the report is focusing on where these changes have occurred.

A: The report focuses on where these changes have occurred. It shows you then how we adjust them. [Sound of pages being turned] So basically, what you do for Aberdeen is that you change the early records at Aberdeen, to make them comparable with the modern record of Dyce. [Sound of pages being turned]

Q: So everything is documented.

A: So everything is documented. [Sound of pages being turned] Let me just find Aberdeen here. [Sound of pages being turned] [Pause] So for Aberdeen, it's got some history of the site, the moves and locations, and it was compared with these places, one is Gordon Castle and one is Edinburgh and another one. So for the period 1871 to 1947, the correction factors were these, degrees Celsius times ten, so this is about 0.7 degrees Celsius cooler, the earlier data were cooled to make them comparable. And so we went through all the stations. So there's Aberdeen, adjustments there. [Sound of pages being turned] [Pause] There's another set of adjustments for a place in Iceland, there's 12 numbers. So where you had reasonable records that were close together, you can actually do this. [Sound of pages being turned] It turns out with a lot of the European sites that a lot of work had been done over the years, the records have been looked at in earlier decades, and so what actually got in to World Weather Records, was pretty good to start with. Anyway, that's all the documents.,

Q: Yeah, so you've got very, very detailed documentation of what's been--,

A: Yeah. And just as an idea of what we'd done [Sound of pages being turned] [Pause] And so F is the total number of stations in different regions here.

Q: So you've got very good coverage in Europe and the Americas.

A: Yeah. And A are the stations that were unadjusted, nothing was done to them. And B is the ones that were homogenised, so you're talking about ten percent of the data roughly. There were a lot of stations that you couldn't examine because the records were too short or there were no adjacent stations. Often those are islands a long way from nowhere, or earlier parts of records that we've just got nothing to compare with. There were a few stations that were incorrect, for various reasons. They've got many of these jumps, or they've got trends relative to their neighbours. So some of those have urbanisation trends, quite a few of those are in the Americas,

Q: Yeah, so it's clear, isn't it?

A: Yeah. And Europe. And E were sites with urbanisation and D is the ones that have these numerous jumps. Yeah, so--,

Q: And are these percentages changing in the modern era?

A: In the modern era, I think because there's a greater density of stations now you can find more problems and they're making improved adjustments now but they're not making too much difference. We did show on the original, this paper from '86, that if you didn't do the adjustments it didn't make much difference to the global average, but it makes a differences to the regional detail. So that's the network you see.

Q: Oh, okay.

A: So, for example, that was then. We had about [sound of pages being turned] 2,500 stations, but obviously stations are different elevations. So stations that are at higher elevations would be cooler than those at lower elevations, and unless you allow for that then you run into a lot

of problems as well. How we got over that is to use temperature anomalies, maybe David mentioned this?

Q: Yes, he did, yeah.

A: This is the '61 to '90 base period. If you take that 30-year average out of stations, then the temperature anomalies are much more contourable across the world. You can include high elevation stations of the Alps, they could still then look similar to stations on the German plain, etc.

Q: And it's very clear from there, isn't it, how few there are--

A: Yeah, but that's actually, this is from the mid-'80s now and it was basically an issue of what had been digitised and what was readily available. We did digitise more and our emphasis was on places that had fewer stations, we didn't put much emphasis into the US and parts of Europe, for example, but into other parts of the world. A lot of countries have digitised more data and made more of it available over the years. So where we had 1,580 usable stations, that means they passed these checks and they had the base period, then we use them. Now, we've got of the order of about 5,000 stations, for both hemispheres, but the southern hemisphere has only got about a quarter of what the north's got because there's less land. [Sound of pages being turned] You've got more data now and you can do improved interpolations, so you can do things like Harry's paper, where we do spatial infilling. But our normal way, the HadCRUT4 and CRUTEM4 data, if there's a gap then there's a gap, whereas Harry's work is infilling that.

Q: Yeah. So those datasets, you're compiling what's actually there, what's recorded?

A: The gridding is just putting each station into its five degree latitude and longitude grid box, so the boxes are bigger in the tropics and they're smaller as you move towards the pole. But also, the variability of temperature increases as you move towards the pole and it also increases as you move to more continental areas than the coast. So when you do the averaging, to produce the global temperature average, it's just a weighted average of all the boxes. And the weights, they're just the cosine of the central latitude which gives more weight to the bigger boxes in the tropics and smaller weights to the ones in the Polar Regions. [Sound of pages being turned] Later work is just extending that, using more data, maybe slight changes to the interpolation scheme. It's not a really complex interpolation scheme, there's many more complex techniques that you can use now, like Kriging and spline interpolation, we're not using any of that for CRUTEM4. So, when you see our temperature maps on the Hadley Centre site, when you see a gap that's because there wasn't a station in that box. So all this was available when all the enquiries came in, but no one was aware of those US Reports because they were done in the '80s. We actually back scanned them and made them available to a House of Commons Committee and so to the other reviews.

Q: So you think it's important that there was degree of documentation and I suppose transparency?

A: Yeah. But it's just that it wasn't readily available online. An issue over the years has always been migrating datasets from different media. The first lot was on magnetic tape, the programmes were on cards, and then as things improved, you had them more online, but you were still using tapes then for the datasets 'cause you didn't have the vast amounts of storage that you do today. But now, storage is no longer a limiting factor it seems, except if you do climate- modelling. So it's much easier to do a lot of this stuff now and you can do a lot more analysis. The other point to make is that we weren't the first to do this, we just thought it was a good thing to do at the time in the mid-'80s.

Q: Oh, okay. What attracted CRU to--,

A: People have done this before. The first person to do this was an Austrian climatologist called Köppen, in the 1870s, 1880s, doing it by hand obviously. Another person who did it, a British guy called Callendar, and he did some work in the '30s and again in the '60s, and he was also doing it by hand. A few years ago I went to a meeting in America, in the early '90s, and sometimes when you go to these big conferences there's lots of parallel sessions. And I went to one at the end of a day on the history of climate, which had a few historians. I met this guy from this college in Maine, who had done lots of work on the Swedish climatologist, Arrhenius, Svante Arrhenius, who was the first person to postulate that putting more carbon dioxide into the atmosphere, then temperatures would warm. So he was looking at his documents and letters and things, from the 1890s, 1900 period. And he said at the end it would be useful if he could access the diaries and records of Guy Stewart Callendar. Well, I went up to him afterwards and said, "We've got those." And he came here a few years later and documented- our archive.

Q: So they were held here at--,

A: They were held here. And the reason we had them is that Callendar had died in '64 and his widow passed them on to a person she knew, she obviously wanted them out of her house or something, space or something. I'm not sure what the rationale was. They were all in shoeboxes and she gave them to a friend. He had done some climate work in the '50s and '60s and he used to go to Royal Met Soc meetings in London. He was a schoolteacher, and he kept them for years, and when he died, his widow contacted us. We've had them here for years, and we had both Callendar's and this other guy, Justin Schove. The book as produced by Jim Fleming, about Callendar. I thought I had a copy of it here, but maybe I haven't, maybe I have taken it home. . Another example is where Rob Allan traced the descendants of another well-known climatologist called Gilbert Walker, who developed something you might have heard of, the Southern Oscillation Index, and had worked in India and Australia and developed various indices of climate around the world in the 1910s, 1920s. A person who used to work here, called Mick Kelly, actually traced Callendar's daughters and went to talk to them. They didn't live too far from here, and one of the daughters remembered they used to talk over dinners and lunches about what the climate would be like when it was much warmer, and they were doing this in the 1950s. So Callendar was quite an innovative person, but he wasn't a scientist in the strict sense. He was an engineer, he worked on fog dispersal in the

Second World War at RAF aerodromes around the country. He was just interested in all this, it was all sort of done in his spare time. He was a steam engineer.

Q: So you had these, there was this history of recording and then you've got the archival records that are there. What was the underlying motivation to get started with the datasets?

A: We just thought it was a good idea at the time.

Q: Did you have any particular goal in mind with it?

A: No. In fact, when we first did it, the first paper came out in '82 and it was just for the land then, the temperatures had been cooling from the 1940s to the late 1970s. When we first did it, it wasn't to show warming, it just to produce a dataset that we could look at. One of the techniques that was used a lot them and still is now, is if you got maps for 100 years was a statistical mathematical technique; in Britain it's called Principal Component Analysis.

Q: Yes, I'm aware of it.

A: And the Americans call it Empirical Orthogonal Functions. So we just wanted to do some analyses of that type.

Q: So it was about working with data.

A: Working with data, and relating that to similar analyses on the pressure data because that was available, to try and see why the variability in temperatures was occurring, and found one of the factors was changes in the atmospheric circulation.

Q: But you didn't really have any particular kind of higher goals in what you would do with those findings?

A: No. What we produced initially, and we have always produced, is gridded temperature databases, and a convenient way of displaying something at the end was to produce the average of it, which was the global temperature average from land. That was a just convenient way of summarising the data, but that wasn't the aim, the aim was the gridded dataset. And like Harry would have told you, lots of people actually use the CRU TS products for the grids. A lot of people in developing countries use the dataset because it's one of the few sources of data they can get access to. A lot of African students come and use it. They want to do something in Nigeria or wherever, but they can't get data from the country, so they can get that we've got.

Q: So has your data always been openly available?

A: The gridded products have always been openly available. A lot of the station data was available in the early 1980s as well, so it's on datasets in the US and we used to update it and we used to send out magnetic tapes to people with the gridded datasets on. Obviously we didn't have the internet then, so we used to send magnetic tapes out around the world, really to fellow scientists.

Q: Was it on a request basis, someone saying, "I need some data."

A: Yeah, but we got the data in to some databases in the US--, you've heard of NCAR, haven't you?

Q: Yeah.

A: It's the National Center for Atmospheric Research in Boulder, Colorado. We got the data there on a magnetic tape, so they were distributing it for us as well. There's something now that NASA runs called the Global Change Master Directory or something, and so the data is there too. Through doing the US work, the Department of Energy set up a big data centre in Oak Ridge, Tennessee, which is called the Carbon Dioxide Information Analysis Centre, and they had the dataset as well, so the data was made available that way as well.

Q: Did you have sharing as a goal or was that something that just occurred because people asked?

A: Just because people asked. In the first couple of papers, we had big tables of the data but it was really a pain to actually proofread these. Then people wanted to have the numbers and so we used to send them round as small files on discs in the past, but the data itself, the station data and the grids, was sent around as well. The change came after the first papers in the 1980s. When we did it again in '94, we realised that a lot more data was being digitised around the world. It was a slow process and so we were trying to make contact with a lot more people. There was a lot more data coming out of Russia at the time as well, and getting some more out of China then, so those were improving areas. Then we managed to get some datasets from some countries, but they said they didn't want the data sent on but they were happy for the gridded products to be sent on. No one really wanted the station data. It's a bit of a mess really, because you've got to deal with all the gaps in the stations and stations at irregular locations. Climate scientists and others are much happier with grids of datasets, you've only got to deal with the odd missing numbers here and there.

Q: What is it that makes the grids more valuable?

A: Well, the series become more complete. So with the individual stations you've got gaps, but if you've got two or three stations in a grid box, then it's liable that each grid box series will be more complete but the individual stations will still have the gaps.

Q: So you're kind of aggregating, and--,

A: Yeah. But in the '90s the issue of data availability became more of an issue. In the '70s and early '80s we just used to ring the Met Office up when they were at Bracknell, ask them for some data and they would send it up on one of those floppy discs or a magnetic tape and there were no questions asked, no one worried about accessing the data. But then the reason that all the data issues started was this. Data was always freely available in the US, what was digitised, so that's what the American Met service, which is called NOAA, had digitised. It was mainly from the post-war period from the late '40s. They have gone back and digitised most things now and they had they made it all available. It developed a big area of consultancy work in the US, not in just in climate but in weather, air pollution and related studies. So there

were a lot of consultants in the US and they were looking to expand their markets. They wanted to come to Europe to do this and a lot of the European met services weren't happy with this.

Q: Why do you think was?

A: Well, they didn't want people coming in and doing work that they thought they should be doing, even though they probably weren't doing it in some cases. So a lot of European countries began to impose restrictions on data because they didn't want the Americans coming over. This has always led to a lot of issues within met services around the world. There's an umbrella organisation for the met services which is called the World Meteorological Organisation in Geneva, but they're just a coordinating body that tries to standardise things and exchanging data between national Met Services--. It's for the members, such as the met services, it's not for others to get access to. I've been on lots of WMO committees over the years and this is always an issue. Eventually they developed WMO Resolution 40, which was supposed to make certain data freely available, but that hasn't really helped because the met services around the world can interpret it in a variety of ways.

Q: So it's not truly standardised?

A: Well, it's not about standardization, it's about availability. So, the monthly data, the monthly average temperatures and monthly total precipitation, was supposed to be freely available, but if the members don't want to make it available then there's no policing by WMO, and they can't make countries do something that they don't want to do. And so what eventually happened is that a few met services start getting data together for more than just their own countries, so the American one is the main one, and the Russian one and Britain. So that's when we began to work more with the Met Office as we had the data that they really wanted.

Q: What particular data did they want from you?

A: They wanted the station and gridded temperatures that we'd had from other parts of the world.

Q: Right, for the non-UK.

A: The non-UK ones. And also in the mid-'80s and into the early '90s, everything had been land only but obviously that's only about 30 percent of the surface, so they were working on the marine data. And that's when the combined datasets were produced, with the sea surface temperature anomalies over the ocean and land-air temperatures over the terrestrial regions.

Q: So that was an initiative from the Met Office?

A: I don't know who the initiative was from, but we were happy to work with them and they were happy to work with us. I used to go down to Bracknell quite regularly.

Q: So how would you describe that relationship, then and now?

A: Oh, it's got better over the years.

[section deleted from published transcript at request of interviewee]

Q: So would you say that was a more sort of unofficial relationship then or more informal?

A: It was more informal. You just used to deal on a scientist-to-scientist level. If you went to ask for British data, it's far easier to go to the scientist, but if you went through the hierarchy at the Met Office, if you don't go to the right person they've got to then find out at their end who has got that data. It's a bit easier now because you can track things a little better, but if you did it formally then there would be always be more issues coming along, particularly with rainfall data, wind data and sunshine data.—

[section deleted from published transcript at request of interviewee]

Q: What sort of other organisations were interested?

A: Well, the Institute of Hydrology was interested in getting all the rain gauge data across the country to look at flooding and changes in rainfall patterns. Then came this big market for certain types of other data, particularly wind data, from where you site wind turbines, and solar, sunshine data. So those two datasets are separate from the others, even though they're all recorded at the same places. So trying to get wind speed data and sunshine data is still quite difficult from the Met Office.

Q: Do they place an economic value on that?

A: Yes, they do. They have a charging policy that if you were doing it for academic research it was fine, provided you didn't pass it onto a third party after the work.

Q: And that's on a basis of trust, is it?

A: Yes, I don't think they policed it that much. The other met services in Europe used to do this. We used to get data from the German met service, which was weird because they used to send you the data and then send you this other letter saying, "If you send this on, we will fine you," and there was this number, about six figures, of German Marks. It wasn't just sort of 100,000, every number was there, they had some formula to work what the fine would be, but they weren't policing it either.. The Germans probably still can't get anything out of Poland, for example.

[section deleted from published transcript at request of interviewee]

Q: Really? Yes. For historical reasons?

A: Well, I don't know whether that's the reason, but the Polish met service have always been awkward over years and you talk to Polish scientists and they know all about it because they have the same problems. Sometimes it's just personalities, I think.

Q: So the temperature data, that's exchanged on a working relationship, is there a cost of that data?

A: No. No, the Met Office are one of the nodes on this global telecommunications system, and they pick it up.

Q: So that's done on a kind of a mutual basis.

A: Since Philip Brohan produced the paper in 2006--, from about the mid-'90s to about 2005/6 they were sending the temperature data and we were doing the gridding and sending them back the grids. But it's much simpler if they just ran the programmes, so we allowed them to do that and they just acknowledge us on their web pages and in their papers as well.

Q: So you supply the data, they then use in various ways.

A: Yeah. It's what David's doing now, he's just improving the network once a year, this time of year. They'll get a new version back and they'll just put that into their software system and pick up a new dataset and produce new grids. The whole dataset will change as bringing in more stations means that some of the earlier data changes as well. But some gaps will still be there, but the gaps are getting less over the years.

Q: After the data has gone back to Met, are you concerned with who uses that data or what products it goes into?

A: No, because all the station data has been made available. We had an issue obviously in 2009, you know, people said, "We need access to this raw data." So between the two of us, we wrote a letter to all the met services, using an address list from the WMO. There were protocols developed for how WMO wrote to National Met Services. These reduced the number of letters sent.

[section deleted from published transript at request of interviewee]

Q: What's the reasoning behind that?

A: I don't know. That was dealt through the members and through the permanent representatives on the WMO for each met service. We'd asked the WMO if they would send out the letter, and they said no they wouldn't, but they said the Met Office can send it out, so the Met Office sent it out. We got the addresses from the WMO, so about 150 letters went out. They all went out in English, normally WMO deals with countries in the six UN languages, but they all went out in English. We didn't expect some countries to reply because they hadn't gone out in Chinese and Arabic and so on, and about 50 countries replied. So the questions were whether we could release the data we had for their country, which wasn't necessarily their data, it was our versions of their data. Most countries were happy with that, well the 50 that replied, obviously 100 didn't reply. But there was a handful of countries that sent back letters saying that we couldn't. Now, we had some discussions with the Met Office. There had been some developments in the networks over the years. There's something that was set up in about 2000 which was called a Regional Baseline Climate Network. This network was supposed to be for data exchange between countries, and it was just monthly averages. It was supposed to be freely available, it was designated as being freely available. So we decided to overrule countries where data came from their RBCN stations. About four or five of the countries that said you can't but they had only one station and it was an RBCN station, so we just ignored them. The only three countries that presented something a little more serious

were Poland, Canada and Sweden. Now Poland said, “You can’t send anything out.” Canada misinterpreted the letter, even though it was in English, and they said, “What’s the rationale for this, why is this so important now?” And you obviously had all this media publicity and they still replied this way.

Q: Yes, of course.

A: So Canada, the Met Office spoke to someone in Canada and sorted that out. What they didn’t realise in Canada was that we were releasing their better data. So if I backtrack a minute. Doing all this homogeneity checking and quality of the data, it became clear that the best people to do that are the met services in a specific country, and a lot of met services have got onto this now, more in developed countries than developing, unfortunately. And so say in Australia, they’ve sorted out most of the data quality issues and made all the data available, and you can even get the unadjusted data from Australia, too, like you can in America. Canada was doing this too. They wanted us to use their adjusted datasets and they said they didn’t want us to send out what Canada was issuing internationally because they wanted us to send out their adjusted data. But what they didn’t realise, they didn’t read the papers that we’d sent with the letter, was that we were issuing the Canadian adjusted data. We were getting that access from someone in Canada through a personal contact. This dataset was on the Canadian website at the time, so that was sorted that. Canada now is still somewhat odd, because you can go onto the Canadian met service website, you can download all their data, but it’s all unadjusted, it’s just the raw data. On another website they’ve got all the adjusted data. There’s no link from the one you first go to, if you’re coming in as a new user, to the adjusted data which is a lot better. There is a link from the adjusted data back to the unadjusted, but there isn't one the other way and they still haven't got this sorted out. I’ve talked to various people in the met service there and Canadian scientists and say they must try and sort this out, but it’s been years and nothing much has happened.

The other country was Sweden. Sweden misinterpreted the letter and they thought we were releasing a lot more Swedish data. We’d sent each country a separate list of what data we had and they misinterpreted it completely and the issue got taken up by a Swedish newspaper, and then the Swedes apologised, because everything was freely available. So it’s only Poland. So we released as much of data as we could in 2010, and in 2011 we decided just to release the Polish data as well.

Q: Did anything happen?

A: Nothing happened.

Q: So now you can get this on the web?

A: Yeah. Quite a lot of the Polish data. So all the data was released in 2011, all the station data.

[section deleted from published transript at request of interviewee]

Q: So why was it important to put the data out there? Was it about the—was it because it was demanded?

A: I think people said the claim made by the enquiries was that the data we use should be freely available.

Q: And that was about transparency?

A: Yeah. But when it comes to other datasets, that's just ignored. Just two examples, quickly. The German weather services have a project to do precipitation data in a similar way to what we've done, and they set up an organisation called the Global Precipitation Climatology Centre within the German Weather Service, and they made contacts with all the met services. And a lot of the met services send them extra data, but they have made a condition that they would not send out the raw station data. So the Germans produce gridded precipitation products of the whole world, different resolutions, but none of the station data are available.

Q: Do you consider that to be a good or a bad thing?

A: Well, no one complains about it because it's precipitation data. But there's grids, you can do the same sort of analysis that we were doing with the PCA [principal component analysis]. The Dutch met service have a project to get all daily data across the European region. It's Russia as well, Turkey and some of the near East countries as well and they've collected all these data from countries. They make a lot of gridded products available and they'll make about 60 percent of the stations available when the countries say that they can make it available, but 40 percent of the station series is not available. That seems to be working fine. They've got the data for Poland, but they are not allowed to make it available except in gridded products and, really, for the scientific purposes the gridded products are good enough.

Q: Okay. So the raw data is put out there because most people have asked for it to be there.

A: I'm not sure of the reason.

Q: Do people use it?

A: Well, it's difficult to tell.

Q: Do you have any kind of feeling for who might use it and what for?

A: I don't know, really, because you can't really check whether people will say what they're doing with it, they've just downloaded it.

Q: So there's no kind of way of checking--?

A: Well, they can count how many people are downloading the datasets.

Q: But there's not any kind of trace of who that is?

A: No. No, and recently a colleague of mine, he's away this week and the next I think, Tim Osborn, partly as a result of the enquiries got some money from something called JISC. Have you heard of JISC?

Q: Yes, I know of JISC, I've been funded by them.

A: So that was to make it more transparent as well, and so we had eight months of money from them to do something, mainly as a result of events in 2009/10. In doing that, they were looking at the issues of the locations of the stations, saying that these aren't accurate enough, which we knew, but for our purposes it didn't matter where in the grid box the station was, as long as it was in that grid box. We've been asking WMO for years to get better locations for data and they've sent round letters over the years to try and get countries to do this, to go to each station and get a better location, and some countries have done that, but not many. So Tim thought another way of making the data more available was through Google Earth and we wrote a paper in one of these new data type journals, *Earth System Science Data*, and it came out a few months ago. And you can download the software and it will come up with a grid over the earth, five degree latitude/longitude grids, and it will show the stations. And you can zoom in and it will show you the time series of all the stations, by seasons, and it will show you the grid box time series too. So that's been made available and people have downloaded that. For this particular journal, it's the most downloaded dataset even though it only came out a few months ago. But we've no idea what anyone's doing with it.

Q: Do you have a feel for who it might be?

A: No.

Q: Do you care?

A: No, I don't really care that much, in terms of what people do with it. We do see odd things where people write papers in obscure journals and say that Station X in our set isn't correct and they then go on to imply that because this one station's incorrect that the whole lot is wrong.

Q: So how do you respond to something like that?

A: Well, in this paper in 2012, maybe I should print you out a couple of papers as well. We knew from back in the '80s that when you've got thousands of stations, you could just take a small subset then you could produce the global temperature average. This was quite accurate with about 100 stations, well located over the land. You got the same answer as you got with 5000, and other people have now found that. So this group in America that started all this again, called the Berkeley Earth Surface Temperature initiative, to get this acronym BEST, a bit contrived. They've taken numerous datasets mainly from within the American met service, within NOAA, NCDC. They did all this and they were funded by some sceptic organisations in the US, and their aim I think to start with was to find out it was wrong. Well, they found out with their technique that it had a warmer trend than ours, slightly. So they then started looking at urbanisation, they thought that was the problem and they then separated out all the stations into urban and rural, based on satellite data. They analysed those and found that the rural ones warmed slightly more than the urban ones but there was no significant difference. There are lots of issues about urbanisation over the years, but people seem to still think urbanisation is why the temperatures are going up. We've done lots of work with the Chinese data recently partly because that's possibly where--, if you can't detect it in China then you're not going to

detect it anywhere, because of the growth of Chinese cities in the eastern part of the country in the last sort of 30, 40 years. But again, it's not an important issue there.

Q: So how do you feel about the sceptics then using your data for their own purposes?

A: Well they can, but a lot of them are misusing. You see things, they just make claims which are just ridiculous on blog sites. We're not looking at blog sites regularly, so we then probably only see a small percentage, just through someone alerting me up to various things.

Q: Do you respond?

A: No. There's no point in responding to blogs. The blogs to me are a complete and utter waste of time. I mean some have some reasonable people on, but they've either got some agenda, and-- they just don't know the scientific literature. It is hard to come into a subject without that background information. What really irks me most is scientists from other disciplines, particularly physics and geology, who come in to climate and try to reinvent things, just because they haven't read the literature. If I was going into another field, I would realise that there's a massive scientific literature there and to try and get up to speed on this. But there's some people who come in, some physicists think they can come in. These Berkeley people were physicists. They come in with an arrogant idea that they're physicists, they can solve everything. And they've proved to themselves that what we've shown is correct. [section deleted from published transcript at request of interviewee] So at least they were reasonable enough to get the same result, but there was only one result.

Q: Who was funding them? Was it corporations or governments?

A: The Koch Foundation. It's two brothers, and their surname is K-o-c-h, so we might call them Koch, but apparently in America they're pronounced "Coke". They've been funding many right-wing organisations all across America in various things and they have input into the Senate and the House of Representatives.

Q: And their interest is economic?

A: They want a status quo, basically, they want to continue burning fossil fuels. So they think also we are rabid greens, but none of us here are members of any of these organisations. We've had a little bit of money from the World Wildlife Fund for Nature (I've forgotten what it's called now) over the years to research one or two things, but when you do work for them--, well, we've also had money from BP and Shell over the years, you give them the results you find. It's the same with any government money or research council money, or EU, you're not giving them what they want, you're giving them what the data says. Whatever the funder, you've written a research project and the project report or scientific paper discusses what's been found.

Q: Okay, so that brings us onto--, you've said that a lot of your work is funded by research councils, other organisations. To what extent does that set your research agenda here, the questions that you're asking or the questions that you're investigating? Well, what I'm trying to

sort of look at then, is do you have your own agenda that's you go out and look for grants in those areas, or do people come to you and say, "Can you research this for us?"

A: Mainly the other way, you're just submitting grant proposals to various calls from research councils and the EU. [section deleted from published transcript at request of interviewee] , We are also trying to keep the people we've got. So you can't just bid for X number of research grants and get them, you have to bid for two or three times X. And so sometimes you get more grants than you've got people and you have to get extra ones; sometimes you don't get enough and so one or two people have had to leave over the years because we've just run out of money for them. But that's the whole issue that you're fully aware with funding research.

[section deleted from published transript at request of interviewee]

Q: I've got one last question about the data and then I'd really kind of just want to ask you a bit more about some of the external organisations, if that's all right.

A: Yeah.

Q: How do you validate or evaluate the work here that you do, how do you say, "This is right, it's good quality?"

A: Well, in terms of the temperature data, there's three groups in America doing something similar, there was a group in Russia in the past and the Chinese and Japanese are doing things now as well, so one of the ways is comparing with them. I know a lot of the stations are in common, but what we're all doing is different in how we all combine the temperature data, how we look at the quality of the data is different. We've even wrote a little--, I'll print off a few papers for you later. Ed Hawkins at Reading realised that Callendar's paper came out in '38, and it was 75 years on in 2013, last year, so he wanted to do a little paper on Callendar's reconstruction in 1938. Callendar actually produced Northern Hemisphere numbers in a table at the end, so he digitised those, only about 100 numbers to digitise, and then we just compared it with what we get now and Callendar's results with about 200 stations are exactly right.

Q: So comparing with the past and with other contemporary work.

A: So comparing with the past--- We know why Callendar's work was right. So what we try and do now is we get this data from more countries. David's going about the quality control of data from the countries, but we don't do any quality control work ourselves apart from what David does. It's not homogeneity, but it's checking what we're putting in. But also in the 2012 paper we did some work where we restricted the network, so we based it on just 100 or 200 stations around the world, and you got the same answer, which the Berkeley people finally realised. They didn't seem to understood why and then tried to write a paper explaining why, without knowing any of the literature. . We also then did something which is partly due to one of the enquiries suggested this, "Why don't you take out loads of the data," and this was what we did. We took out all the American data. So you take out all the data from the 48 contiguous US

States, you leave Alaska in and Hawaii, and there's no difference until you get back into the 19th century. And we did the same with taking out all the Chinese data, all the Russian data. You do get an effect if you take out the Soviet Union as it starts making a bit of a difference then because you are taking out about 20 percent of the surface area of the earth. But the US is only about 2.5 percent of the surface area of the earth. We did an average test where you take out all the Australian data from the southern hemisphere, and that's about half the data, but in terms of the grids, it makes little difference whatsoever. The whole thing is quite robust. Coming back to reanalysis, you know what reanalyses are, basically?

Q: Yes, I have a lay person's understanding.

A: Over the years, weather forecasts have improved and that's because you're getting more data in. Since the '70s they're getting a lot of satellite data and they've had radiosondes, weather balloons, and the computers have improved dramatically so they're able to reduce the resolution, so weather forecasting has improved dramatically over the years. The sceptics still don't believe this, and there's a little aside at the end om this. So what's done in a reanalysis is that you've got a consistent climate model and an assimilation scheme, and you just put the data in. Some reanalyses include all the data as it comes in over the years, so the balloons come in with the post-war period, the satellites come in in the late '70s. But another way of doing it is to just put in surface pressure data only, measured at stations around the world and on ships, and that's what extended reanalysis produce that go back to the 19th century do. So they're not even putting temperature data in, apart from sea temperatures. All weather models have sea temperatures going into them. It's not coupled so they can't interact with the ocean, they just tell you the ocean surface conditions underlying the model. So all this goes in and it produces temperatures everywhere, and those temperatures are pretty much the same as what you'd get if you'd put the raw temperatures in. So by just putting sea temperatures, sea ice boundaries and surface pressure data we recreate the global temperature record.

Q: I see. And so the reanalysis work you're getting involved in here now, is that--, or are you feeding into it?

A: We're feeding into that, it's weather data rescue. As you go further back in time, obviously networks get sparser, obviously there's big issues with the Antarctic before the International Geophysical Year, but it doesn't seem to have much of an impact on the rest of the world. But you could, I mean Philip Brohan's plan is to get these reanalyses back to 1800; they go back to 1870 now. But if you can get the sea temperatures you are part of the way there, but the sea ice distribution is a real problem. You can make assumptions about the quality of the sea temperature data back in time and the sea ice boundaries, and assess that, run the reanalysis with different series, boundaries, and you're probably going to get pretty much the same thing.

Q: So do you see this being an important area of work for CRU in the future?

A: Yeah, but I think it's more on analysing the results as we're not going to run the reanalyses. We will help with some digitising efforts. You know, Clive finding all these ship logbooks is one-- And a lot of these, as Clive said, a lot of them are British ships.

Q: What's driving that? Is it the work of the Met Office or are there other factors that are--,

A: Other groups around the world. It's just that vast amounts of papers now are based on analyses, of reanalysed output. A lot of people just refer to them as observations, but they're not. Observations are going in, but it's an assimilation system. And in terms of weather forecasts, some of these reanalyses are amazing, going back in time. You know, I went to a meeting a year or so ago and a Portuguese guy got up and said that they looked at the reanalysis and they found this tremendous rainstorm near Lisbon in about 1880 and they wondered whether this was real. They thought they'd found a problem with the reanalysis. But they went back, and on the two or three days when this storm hit southern Portugal and produced this massive amount of rain, they found it in the newspapers and there was the storm. They went back and they found the data and digitised the data, and there was the storm. They just hadn't digitised the data back far enough, the storm actually took place, the flood marks are on the buildings, on the Tegus River upstream of Lisbon, they'd just forgotten about it basically.

A: So that historical data is becoming more important in climate science?

A: Yeah, and all that went in to get that was a few pressure readings from Iberia, the Azores and the Canary Islands, and the system produced it. And, in terms of weather forecasts -- the sceptics say the weather forecasting is poor, but it's improved dramatically over the years. There are two weather forecasting systems that are the best in the world. The second one is the Met Office. It's better than anything in America, or Germany, or anywhere else. But the best one is also in Britain. It's one you've never heard of probably, it's called the European Centre for Medium Range Weather Forecasts which is at Reading.

Q: Ah now, I have heard of it.

A: It is at Reading. So that's an organisation set up in the late '60s and '70s, to coordinate European met services. I'm unsure why it's located in Britain, but it was very close to the Met Office when it was at Bracknell. But they've had a concentrated effort and they've got into reanalyses now in the last few years. Their model is much better than the American models and some of their reanalyses are just amazingly good. And the Met Office are trying to compete with them. And if you look at real-time weather forecasts and how they're assessed internationally, these are the top two models. US models are down at number five and six because they've not put the efforts into research in that area.

Q: So is the value in that, is it the data, is it the model, what's the best--, ?

A: It's the model.

Q: It's the model itself--?

A: It's the assimilation and modelling system that they've developed at the European Centre. And the Met Office are very close they're also doing very well. There's a lot of interplay between the people involved. This weather forecasting system is only good because of the data that goes in. Take the data away, it wouldn't be any good. I heard a story that in New

Zealand in the 1990s there was some economic trouble and they were trying to cut back. So the Minister went to the New Zealand met service, saw the forecasting system and saw what they did, and they explained to him what they did. And New Zealand isn't a big country and they haven't got the resources to run their own model, so they were picking up their weather forecasts, which are sent round internationally, from the European Centre, from the Met Office, from America, from NOAA and from Australia. And they got these in, and they built up experience over the years that for certain types of weather systems, they know that the British model or the European model works better, or the Australian model may work better in certain instances. So the Minister took from that, that the New Zealand met service didn't need all the measurements they were taking, so the radiosondes that were going up from several places in New Zealand, a lot of the surface measurements weren't needed, because they were getting all the weather forecasts from these other countries, other systems. And so he decided, a few days after this meeting, to stop funding the launching of radiosondes across New Zealand. And the New Zealand met service then had to go back and explain that if they didn't have that data going in, the forecast would be much poorer. So he completely misunderstood. And that story is told time and time again. So when governments discuss different met services, even if they're using the European Centre forecasts across Europe, the importance of the observations that the met service in that country is taking go into the system to provide the good forecasts.

Q: So the very start of it is, you know--

A: It's the data that goes in. If the data dropped out, these forecast systems would go downhill quite quickly. The surface network is less important, but the radiosondes are the key thing. [section deleted from published transcript at request of interviewee]

Q: And there are relatively few of those?

A: There's about 800 of them, but there's only about 500 or 600 that do it regularly. And you can find out that if you took one radiosonde out it degrades the forecasts over much of the tropics. There's one that goes up from the Galapagos Islands over the El Nino region, and if that disappears, which it has done over the years because Ecuador hasn't got the resources, then it degrades the forecast most of all; taking one type series out. So they've done lots of experiments like this and shown this, it's in a number of scientific papers.

Q: So you're quite sensitive to those resources being available.

A: Yeah. So you could actually find these things out by looking at the data. And that's what the European Centre are doing now, and what the other reanalysis centres are doing in Japan and America, is trying to--, it's been quite difficult to actually show this. So what they're going to produce now with the latest reanalysis that the European Centre is doing is to release all the input observational data in a feedback archive, in a big EU funded project just finished [section deleted from published transcript at request of interviewee]

Q: Right, so you're kind of unpacking the data.

A: Yeah. So it's in an assimilation system, but you're trying to find out what the key observations were they put in . And it'll also determine what the key satellites are.

Q: And that will allow you to build better models?

A: Possibly, but it will allow you to emphasise which datasets are really important. So it might tell you if you had an extra radiosonde here or there where it's going to make the biggest impact. It'll also tell you which satellites are doing the best things as well.

Q: So yeah, optimising their use.

A: Yeah. So it's been quite difficult to do and they still think it's going to be very difficult for people to interpret, because the files will be so large.

Q: So it strikes me that there is some kind of need for international level--

A: Oh, yeah, it's all going on through WMO.

Q: So they're quite crucial in making sure that--

A: There's a group within WMO called GCOS, which is the Global Climate Observing System, and they're trying to maintain networks. So they have some money from donor countries, mainly European and America, to try and make sure measurements are made in certain parts of the world. So they're actually paying Ecuador money to put these sondes up.

Q: Because it needs to be done.

A: Yeah. And they're trying to get a few countries in Africa to do a few more. But it gets a bit political then because they won't fund certain countries, but they'll fund others. It's just a sphere of influence issue for the donor countries.

Q: So who decides that? Is it the members of the WMO or--,

A: Well, there's the group that I'm on that makes recommendations where to put their resources, but it depends if the country wants to pick up the tab for certain things. So the Met Office recently picked up an issue, and I don't know why they did this, but the network had degraded totally in Madagascar, so they've set up a whole system to monitor automatic weather stations everywhere together with a telecommunication system to the Centre, and then to send the observations out. They're all working through the mobile phone system, doing the communications from the outstations to the Centre in the capital. And then the government in Madagascar changed the mobile phone provider for the country, and the system just collapsed and so they had to then send out some engineers to get it to work again.

Q: So yeah, it's quite fragile in countries where there are less robust systems in place.

A: Yeah. There's quite a few interesting asides of mobile phone systems being used for this and also mobile phone systems, the signals. The companies have noticed this too, that they're distorted by weather, in particular they're quite a good guide to measuring heavy rainfall events because that's what distorts signals the most so they're actually using information from these now. So also what goes in now is other things like commercial aircraft monitoring, not

just the temperatures and humidity and things, but concentrations of gases and aerosols. So they're trying to get the key airlines to have them on their systems and they're really keen to get them on some southern hemisphere airlines, like Qantas,

Q: So do you see commercial organisations becoming more involved in the whole kind of data system?

A: Yeah. And if you get the reasonable ones and they understand--, Lufthansa has some and I think some are on British Airways flights too, but on some more commercial airlines in America and Europe it's harder to get them to do this. And I mentioned the ships, ships have done this for years and years over the oceans, but they some shipping companies are observing less. Because they launch lots of things for scientists too, such as buoys that are dropped, drifting buoys that improve weather forecasting in the southern hemisphere, but some are less keen to do that. So it's hard to maintain the network. You put the drifting buoy out, it doesn't stay there for long and it only lasts for about a year, so you've got to continually put these things out.

Q: Right. So what's holding them back? Is it finances, is it the environment?

A: It's probably finances and some of them they don't really want the people to know where they are and what they're doing for commercial reasons, and then there's security reasons in some parts of the world. This has been the case for years because there's fishing fleet data that goes in two or three years late, all the time, and has been done for years and years. But it goes in, it just doesn't get used for weather forecasting but it gets into the reanalysis. And the fishing fleets from Japan, South Korea and Taiwan are trawling the whole South Pacific, taking lots of SST measurements. But they don't want that to go in in real time because they will then tell the others where they are when they're all trying to get the last tuna out from somewhere.

Q: But they're benefitting from it in some kind of way--

A: Yeah. The whole aim of the original network, and Clive may have mentioned this was that ships take these measurements. The whole idea in a conference in 1853 was that the ship logbook data would eventually lead to improved weather forecasts and less loss of shipping in storms. And it's eventually happened, it's taken 150 years, but it has happened. The one issue now people complain about, still, is that ships now have such good weather forecasts that they avoid the storms. And so some people believe that there's less storminess because ships are avoiding the storms now.

Q: Okay. Great. Thank you.

A: Okay. So I'll just give you a quick tour around and show you a few things downstairs. If you then want to sort of go through things and have a check of where you are, and if you feel you've missed something, then come back later before you--,

Q: Okay.

[END OF INTERVIEW]